DOCUMENT RESUME

ED 174 684 TM 009 656

AUTHOR Crocker, Linda; And Others
TITLE Predicting Children's Writing Performance from Standardized Achievement Tests.
PUB DATE Apr 79
NOTE 14p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Composition Skills (Literary); *Creative Writing; Essay Tests; Expository Writing; Grade 4; Intermediate Grades; Objective Tests; Performance Factors; *Predictive Ability (Testing); Reading Comprehension; Scoring; *Standardized Tests; Test Reliability; *Writing Skills
IDENTIFIERS Holistic Evaluation; Interrater Reliability; *Metropolitan Achievement Tests

ABSTRACT

The relationship between children's performance on a standardized achievement battery and compositional writing performance was examined. One hundred thirty-eight writing samples were collected from fourth-grade students. Compositions were scored for both mechanistic and holistic qualities. Four subscores on the Metropolitan Achievement Test were also obtained: word knowledge, reading comprehension, language arts, and spelling. Canonical correlation and multiple regression were used to determine the degree of relationship between these variables. Results indicated that stanadardized test scores have moderate predictive value for children's holistic writing performance. (Characteristics of children with differential performance on writing exercises and standardized tests were also examined, and several implications for teachers were discussed). (Author/MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174684

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Predicting Children's Writing Performance from Standardized Achievement Tests

Linda Crocker Toni Ondrasik Linda Lamme

Institute for Development of Human Resources

University of Florida

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Linda Crocker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM009 656

Presented at the Annual Meeting of the
National Council on Measurement in Education,
San Francisco, April, 1979

ABSTRACT

The relationship between children's performance on a standardized achievement battery and compositional writing performance was examined in this empirical study. One hundred thirty-eight writing samples were collected from fourth-grade students. Compositions were scored for both mechanistic and holistic qualities. Four subscores on the Metropolitan Achievement Test were also obtained. Canonical correlation and multiple regression were used to determine the degree of relationship between these variables. Results indicated that standardized test scores have moderate predictive value for children's holistic writing performance. Characteristics of children with differential performance on writing exercises and standardized tests were also examined.

Introduction

As American educators once again focus their attention on the "three R's", measurement practitioners are expected to provide appropriate procedures for assessing student achievement in basic skill areas. The assessment of writing skills presents some particularly challenging problems and raises issues which were once considered "laid to rest". Until recently, measurement authorities maintained that writing ability could be assessed quite satisfactorily through use of objective testing procedures. Today, however, the sentiments of many specialists in measurement and language arts have been expressed by McCaig (1977) who stated, "to evaluate achievement in writing, evaluate the writing of children". The inclusion of writing samples by the National Assessment of Educational Progress and the College Entrance Examination Board in their testing programs underscores the popularity of this position.

Nevertheless, obtaining reliable estimates of writing ability continues to be an expensive, time-consuming process (Coffman, 1971). A number of studies have demonstrated that writing samples must be obtained from students in multiple modes on multiple occasions and scored by several raters to obtain estimates of students' true abilities (McColly, 1970; Braddock et.al., 1963; and Llabre, 1978). Although it may be necessary to evaluate actual student writing samples for certification of mastery or diagnosis of specific difficulties, it still seems advisable to determine if standardized

achievement test scores can be used at all for predicting writing performance of young children. If there is a substantial relationship, teachers could then be encouraged to use standardized test information for such purposes as preliminary grouping for instruction in writing. Also if, as many teachers claim, there are children whose writing abilities cannot be adequately predicted from tests, it would be useful to know if these children share common, identifiable characteristics.

Purpose of this Study

This study was designed to explore the following questions:

(1) To what extent can writing performance of young children be predicted from scores on a standardized achievement battery?

(2) What are the characteristics of children who write well, but who do poorly on standardized tests and vice versa?

It is important to note that although many investigators have explored the relationship between objective tests and essay tests or between objective writing tests and writing performance, most of these studies involved students at the secondary or college level (e.g. Noyes, 1963; Klein and Hart, 1968, or Godshalk, et.al., 1966). Few researchers, if any have explored these questions for young children using general achievement batteries or applying multivariate analytic techniques used in this study.

Methodology

Sample.

A total of 112 children from four fourth-grade classrooms participated in the study. The student sample was evenly divided between male and female and was comprised of approximately 70% white and 30% black children.

Instruments.

Standardized Test Battery. The Metropolitan Achievement Test (elementary form) was used in this study. Four specific subtest scores were used: Word Knowledge, Reading Comprehension, Language Arts, and Spelling. Although the subtests in Language Arts and Spelling would normally be considered the most closely related to writing ability, scores on the reading subtests were included in the hope that they might provide insight into differential performance by some students on standardized tests and writing exercises. The general achievement battery was chosen instead of a specific objective test of writing ability because data from such tests are more generally available to classroom teachers.

Writing Exercises. Two different types of writing samples were obtained for this study. For the first writing assignment, the children were required to write a fictional short story to explain some phenomenon such as "How the Camel Got His Hump" or "How the Elephant Got His Trunk", etc.. Children were permitted to select one of several animals as the topic for the stories. For the second writing assignment, the children were required to prepare a report from a list of facts on the topic of "Bats". Two different types of writing assignments were used so that the generalizability of the relationship between test scores and writing performance could be tested across assignments. These types of assignments were chosen because they typify the kinds of assignments commonly used to teach compositional writing in the classroom.

Scoring the Writing Samples. Writing samples were scored for a total of eight mechanistic and holistic qualities. Mechanistic subscores for each composition included: number of words; number of T-units (thought units); spelling errors; capitalization errors; punctuation errors; mechanical errors (e.g. runon sentences); and word usage errors. To equate for the fact that different children wrote themes of different lengths, number of T-units and number of errors were adjusted to number per 100 words. In addition, four raters rated each composition for holistic quality using a 7-point scale. The average of the four ratings was used as the measure of holistic quality. Inter-rater reliability estimated by analysis of variance (treating raters and topics as ran om factors) was .70. Inter-rater reliabilities between rater pairs ranged from .65 to .77, with a median value of .72. This is comparable to reliabilities reported by Veal and Biesbrock (1971) in their standardized scale for rating essays of students in primary grades.

Data Collection Procedures.

The standardized test was administered as a regular part of the 1977 county-wide standardized testing program. In fall, 1978, writing samples were collected in the regular classroom setting but the instructions were given by one of the investigators with the teachers in attendance to insure standard conditions of administration. Children were informed that the writing samples would be graded and returned to them by their teachers so that they would be motivated to perform as well as they would for any other writing assignment. One week elapsed between the two writing assignments to avoid affects of fatigue and boredom.

Analysis and Results

The first question addressed in this study dealt with the extent to which standardized test scores were related to writing performance. Both multivariate and univariate analytic procedures were used to explore this area of inquiry. First, a multivariate procedure, canonical correlation, was used to determine the amount of common variance between the set of four standardized test scores and the set of eight measures of writing proficiency. This seemed especially appropriate because most teachers consider "writing proficiency" to be a combination of mechanistic skills and holistic quality which occur simultaneously in a given sample of student written work. Results of the canonical variate analyses are presented in Tables 1 and 2. Separate analyses were conducted for the fiction and factual report compositions to illustrate the generalizability of findings across different types of written assignments. Results of the canonical correlation analyses can be summarized as follows:

(1) For the fiction writing assignment, one canonical variate was extracted which accounted for a significant proportion of variance in the two sets of variables. The canonical correlation of .70 was significant at $\alpha = .05$.

(2) The standardized test variable with the greatest weight on the canoncial variate was Reading Comprehension. The writing proficiency variable with the greatest weight was the Holistic Rating.

(3) For the factual report assignment, the canonical correlation (.64) between the two sets of variables fell short of the .05 level of significance ($p = .07$). In view of this, examination of the variable weights is questionable (although again the major weighting among the the standardized test variables was for Reading Comprehension and on writing proficiency, the Holistic Rating received the major weighting.)

From the canonical analysis, the holistic rating was identified as the most important of the eight measures of writing proficiency. To determine the degree to which the four standardized test scores were related to holistic writing the four test scores were regressed on the single criterion of the holistic rating for the fiction assignment. (This follow-up analysis was performed only for the fiction assignment because the canonical correlation for the factual report assignment was not statistically significant.) The correlation matrix for the five variables entered into the regression analysis is presented in Table 3. Results of the multiple regression analysis are presented in Table 4. The multiple R of .55 between standardized test scores and holistic writing score was significant at α = .05. Examination of the standardized regression coefficients for the four tests and their corresponding F-ratios indicated that Reading Comprehension made the significant contribution to the multiple correlation. The failure of the other three subtests to receive significant weights in the regression equation does not mean that these variables were unrelated to holistic writing proficiency. It simply indicates that they did not share additional unique variance with the criterion beyond that partialled out with the Reading Comprehension subtest.

A second focus of this study was the children for whom there were discrepancies between objective test scores and writing performance. Scores on all measures were converted to standard z-scores. A discrepancy in performance was defined as a difference of ± .5 z-score points between holistic rating of writing proficiency and the Reading Comprehension usbtest. Thirty-seven students displayed such performance discrepancies (15 students performed better on the objective test than on the writing exercise; 22 performed better on the writing exercises.) No consistent relationship was observed between demographic characteristics of race or sex and performance discrepancies.

Discussion

For the fourth graders who participated in this study there was a relationship between standardized achievement scores and writing proficiency. The somewhat unexpected finding was that the Reading Comprehension subtest surpassed the Language Arts subtest as a predictor of writing performance. This seems more reasonable, however, when the tasks involved in each subtest are analyzed. In the Reading Comprehension test examinees are required to read a paragraph and identify main ideas or draw inferences from the passage. In holistic rating of a composition, evaluation is based on the writer's ability to organize and express his thoughts around main ideas in logical sequence. In short, reading comprehension requires recognition of main ideas in written passages, while effective composition requires production of main ideas and their elaboration within the context of written passages. Thus the student who has failed to acquire the concept of what a "main idea" is or how it functions, probably can neither read with understanding nor write with clarity. By contrast in the Language Arts subtest the examinee must read sentences and identify errors in punctuation, capitalization,etc. Thus performance on this subtest would be more closely related to mechanistic scores on a composition than to the holistic rating of composition quality.

When test scores were related to holistic ratings on factual reports, the relationship was not significant at the .05 level. The failure of this second canonical correlation to achieve statistical significance underscores the need to assess student writing in a variety of modes to determine generalizability of findings. From a less technical viewpoint, however, the relatively high canonical r-value (just missing the .05 level of significance) and the similarities in the loading patterns on the first canonical variate for the fiction and factual writing samples seem to support, rather than contradict the first finding.

In any event, the results of this study must be interpreted with caution. Although the study offers some evidence of overlap between the abilities that underly standardized test performance and the abilities that underly writing performance, the strength of the relationship observed was insufficient to suggest that standardized tests can be used to replace actual measures of writing. Furthermore, any attempt to establish predictive equations using the regression weights derived for the sample in this study should be especially discouraged since no cross-validation has been conducted.

What then, if any, are the implications of this study for assessment practices at the classroom level? Viewing the outcomes of this study within the broader context of research in the area of writing assessment, the following recommendations are offered:

1) When grouping students for instruction in compositional writing, teachers may wish to use both standardized test scores and actual writing samples as corroborating pieces of information.

2) More than one types of writing sample should be considered.

3) Teachers should not consider Language Arts or Spelling subtests as the only subtests related to writing performance. They should also attend to students scores in Reading Comprehension, since this has been shown to have significant relationship to writing.

4) When test score and teacher rating of writing yield conflicting information, the teacher may wish to review the composition to check on the quality of student handwriting. Has poor handwriting resulted in a lower estimation of the theme's holistic quality? Chase (1968 and 1979) has noted the effects of handwriting on ratings of essays.

5) Teachers should expect some substantial discrepancies to occur between test scores and writing measures. (In this study more than half of the students displayed performance discrepancies of at least half a standard deviation.)

6) If both theme ratings and standardized test scores are used for placement, the themes should be grades prior to reviewing the test scores. Chase (1979) found that expectancies of student performance can influence raters' essay grades. It seems likely that standardized test scores could create such expectations about student performance.

References

Braddock, R., Lloyd-Jones, R., & Schoer, L. Research in Written Composition. Champaign, Ill.: National Council of Teachers of English, 1963.

Chase, C.I. The impact of achievement expectations and handwriting quality on scoring essay tests. Journal of Educational Measurement, 1979, 16, 39-42.

Chase, C.I. The impact of some obvious variables on essay test scores. Journal of Educational Measurement, 1968, 5, 315-318.

Coffman, W.E. Essay examinations. In R.L. Thorndike (Ed.), Educational Measurement, Wash., D.C.: American Council on Education, 1971.

Godshalk, F., Swineford, F., and Coffman, W. The Measurement of Writing Ability. New York: College Entrance Examination Board, 1966.

Klein, S.P. and Hart, F.M. Chance and systematic factors affecting essay grades. Journal of Educational Measurement, 1968, 5, 197-206.

Llabre, M.M. An application of generalizability theory to the assessment of writing ability. Unpublished doctoral dissertation, University of Florida, Gainesville, Fl., 1978.

McCaig, R.A. What your director of instruction needs to know about standardized English tests. Language Arts, 1977, 54, 491-495.

McColly, W. What does educational research say about the judging of writing ability? Journal of Educational Research, 1970. 64, 148-156.

Noyes, E.S. Essay and objective tests in English. College Board Review, 1963, 49, 7-10.

Veal, L.R. and Biesbrock, E.F. Primary essay tests. Journal of Educational Measurement, 1971, 8, 45-46.

Table 1. Results of Canonical Correlation between Standardized Test Scores and Measures of Writing Proficiency in Fiction. (N = 70)

Variables	Standardized Coefficients First Canonical Variate	Canonical Corr.
Standardized Tests		
1. Word Knowledge	.25	.70 *
2. Reading Comprehension	.64	
3. Language Arts	-.23	
4. Spelling	.37	
Measures of Writing Proficiency		
1. No. of Words	-.44	
2. No. of T-Units	.17	
3. Spelling Errors	-.29	
4. Capitalization Errors	.04	
5. Punctuation Errors	-.14	
6. Mechanical Errors	-.34	
7. Word Usage Errors	-.14	
8. Holistic Rating	.88.	

*-- Significant at $\alpha = .05$
(Wilk's $\Lambda = .3807$, $\chi^2 = 60.36$, with 32 df, $p \leq .002$)

Table 2. Results of Canonical Correlation between Standardized Test Scores and Measures of Writing Proficiency in Factual Report. (N=68)

Variables	Standardized Coefficients First Canonical Variate	Canonical Corr.
Standardized Tests		
1. Word Knowledge	-.47	.64
2. Reading Comprehension	1.14	
3. Language Arts	.09	
4. Spelling	.21	
Measures of Writing Proficiency		
1. No. of Words	-.20	
2. No. of T-Units	-.02	
3. Spelling Errors	-.12	
4. Capitalization Errors	-.44	
5. Punctuation Errors	-.22	
6. Mechanical Errors	-.43	
7. Word Usage Errors	-.22	
8. Holistic Rating	.64	

(Wilk's $\Lambda = .4833$, $\chi^2 = 43.99$, with 32 df, $p \leq .07$.)

Table 3. Correlation Matrix for Standardized Test Scores and Holistic Rating

Variables	1	2	3	4	5
1. Word Knowledge	1.00	.84	.58	.59	.45
2. Reading Comprehension		1.00	.58	.59	.52
3. Language Arts			1.00	.45	.30
4. Spelling				1.00	.43
5. Holistic Rating					1.00

Table 4. Results of Multiple Regression of Standardized Test Scores on Holistic Rating

Multiple R .55

F with 4,65 d.f. 6.972*

Variable	Standardized Regression Coeff.	Standard Error of Regression Coeff.	F
Word Knowledge	-.0034	.1951	0.000
Reading Comprehension	.5005	.2240	4.989*
Language Arts	-.1281	.1455	0.776
Spelling	.1845	.1339	1.898

*-- Significant at alpha= .05